Machine Learning Concept Using K-NN Algorithm for Heart Disease Discovery and Drug Prescription

Sanjay Murali[1] & Vivek Chanddru[2] & Shrinaveen S[3]
Department of Computer Science and Engineering
R.M.K. Engineering College, Kavaraipettai, India

***Abstract* - we are living in a world where it is very critical for a physician to predict the disease a person is suffering from and to suggest medicine based on the diagnosis. Evidence based medicine (EBM) is the accurate use of evidence available in making decisions to improve the care given for individual patients. Every year, a significant number of researches (potentially serving as evidence) are reported. With the help of this burgeoning amount of health records from tens and thousands of patients, researchers and medics are not just challenged to leverage the full potential of EBM but also to use it to treat their patients for the best results. We develop a system which would analyse the patient's health with the help of a series of questions and determine the disease based on the evidences gathered previously. We also aim to provide the user with the best known medicine for that particular disease (Also based on EBM)**

***Index Terms* - Big Data, Machine Learning, EBM (Evidence Based Medicine), Hadoop**

I. INTRODUCTION

Most of the developed and developing countries are spending a huge percentage of their GDP on healthcare and for the development of healthcare sector. However, such spending has not yet been successfully translated into quality of care, this can be seen when a large number of patients die annually in hospital because of medical and diagnostic errors. Moreover, there is a significant gap between the healthcare we aim to have and the health care that is currently available. One of the highly sought areas for improving the healthcare sector is the usage of big data analytics (BI&A) techniques to collect, analyze, curate, and present evidence at the point of care, i.e., the practice of evidence based medicine (EBM). EBM means affirming an individual clinical expertise with the best available external clinical evidences.

In our project, we implement Big Data analysis along with the Evidence Based Medicine (EBM) to correlate the symptoms of a particular patient obtained with the help of a sequence of questions to the most probable disease and also find the medicine for that particular disease with the highest success rate based on the previous statistics gathered. The proposed system specifies the use of Big Data and EBM to efficiently determine the disease along with which, in our modified system, we provide a new feature to determine the drug which have been more successful in treating a particular ailment.

II. SYSTEM STUDY

The system study phase analyses the demerits of the existing system, defines the objectives to be achieved by the new solution and evaluates other solution alternatives. Thus it helps in providing insights on how the present system works and can in turn help us improve it by suggesting alternative implementations.

A. Existing System

In the existing system, normal Data Mining based Disease Learning Analysis are available for a structured data. But most of the data available for us are in the form of unstructured data. There is no Evidence Based Medicine Analysis. Big Data Analysis is not utilized to its fullest potential. Data Mining is used for a smaller volume of data analysis. In case of handling huge data, time consumed is high and the cost incurred in processing such huge amount of data is also very high. Data mining in this system can only manage structured data and not unstructured data.

Other Drawbacks of Existing System:

1. It cannot manage large set of data and unstructured data
2. It is not focused on Drug Discovery Process.
3. Lacks Quality of Service.
4. No automatic detection of disease

B. Proposed System

In the proposed system, Evidence Based Medicine Analysis is achieved using Big Data Technique. This Process is achieved by the following stages:

1. Analysis of patient's health condition which deals with all the bio medical reports of the patient. All these reports are processed for analysis. Bio medical investigations include factors like Blood Sugar, Lipid profile, Height, Weight, BMI, BP. It also takes into consideration heart Problems and other problems which are person specific.
2. Formulating questions to the patient will yield prospective analysis of disease pattern, like whether the source is hereditary, tropical region or some other pattern.

3. Evidence Gathering & Analysis will gather all the information like symptoms, bio-medical investigation reports and disease pattern together for Big Data Analysis. We split the data into smaller parts and store in separate data nodes for easy access and to save time while processing.

4. Resultant Output Module will yield the result of disease based drug discovery. In the Modification part, we modify the proposed technique to add an automatic machine technique for Medicine Analysis is of no use. We are implementing the Proposed System together with the technique of Machine Learning based disease discovery. Until we discover the exact disease it is useless to go with drug discovery.

III. SYSTEM ANALYSIS

Systems are created to solve problems. One can think of the system approach in an organized way of dealing with a problem. This also includes looking at the existing system to see what it is doing for the organization and how well the system is doing its job. The feasibility of the project is also considered.

1. Problem Statement

The Problem statement helps in identifying the underlying problem in the existing system and helps finding ways to overcome the same by analysing the problem thoroughly.

A. Problem Identification

We mainly focus on the efficient use of medical records and other personal traits to easily establish a relationship and determine accurately the disease which the person most probably has and to prescribe drugs which are efficient in treating the ailment. Research states that there are many deaths all over the world due to improper diagnosis and medicine error. This causes a serious fall for the healthcare industry and a fear among the people about the quality of the healthcare offered in their locality. There are many countries which do not have proper doctors and those people cannot afford to travel longer distances for treatment. We aim to provide a simple user interface which tries to improve the quality of the healthcare domain.

B. Problem Analysis

The most common solution for the problem would be to contact the doctor available virtually. But this could not be possible in all cases. Also, people cannot always afford to spend a lot of time and money looking for expert opinion in most of the cases. The most probable solution would be to create an online environment, which could collect the medical records of the patients, their family traits, their other personal and environmental factors through a series of questionnaire. Once the data is collected, Evidence based search is done where the system looks for similar symptoms and deciding patterns. The situation with most probability will be selected as the best match and the system declares that the patient is most likely to suffer because of that particular ailment. Once the disease is found, the system then uses Map-Reduce algorithm to find out the best drug for treating that particular disease.

2. Scope

A.Product Scope

The product scope gives us a holistic view of the product. The product scope defines what the product will look like, how will it work, its features, etc. Our system will play a major role in improving the quality of the healthcare system and it could be available for users throughout the world and with the help of professionals, the accuracy of the system could be improved and made available for use by patients easily.

B. Business Scope

Business scope contains Textual Scope Description, list of external actors and their roles, description of business activities. The actors of the system are the user who uses the system and the server which hosts the system itself. The users could be the potential patients or the patients who needs suggestions regarding their disease. The server on the other hand, stores all the evidences and data regarding the users and use those data to fetch the most probable disease which could be affecting the person's health.

3. Resources Required

Resources are required to carry out the project tasks. They can be people, equipment, facilities, funding required for the completion of a project activity. The system would require the knowledge of professionals, the users could also be considered as knowledge providers in addition to their normal routine of being knowledge seekers. The development environment required should have a minimum configuration of Pentium IV processor, with 512MB of RAM and 80GB of memory space. The software resources include the operation system, Hadoop and other editors.

4. Feasibility Study

The feasibility study is to determine whether the solution is achievable, given the organizations, resources and constrains .By performing feasibility study the scope of the system will be defined completely. Most computer systems are developed to satisfy a known user requirement. Once the decision is made a report is forwarded and is known as feasibility report. Feasibility study is an evaluation of system regarding to its workability, impact on the organizations, ability to meet user needs and effective use of resources

The feasibility studies are under 3 contexts

i. Technical Feasibility
ii. Economic Feasibility
iii. Operational Feasibility

i. Technical Feasibility

In the proposed system technical feasibility centers around the hardware and the software and to what extent it can support the proposed system. The tools are used to develop the application are the best tools available in the technological scenario and hence it requires efficient and versatile programmers with good programming skills. Even though the technical requirements are needed for development of the system, any organization which has a web browser and internet connectivity could easily make use of our system. Hence the proposed system is technically feasible.

ii. Economic Feasibility

Economic Feasibility is used for evaluating the effectiveness of the system. The procedure is used for determining the cost and the benefits or the savings that are expected from the system and compare with the cost. In our project all the required facilities, both hardware and software to be used, initially may prove to be costly, but when put to use it proves to be much more economical than that of the existing system. Regarding the maintenance, since the source code will be with us, any small necessary changes can be done with minimum maintenance cost involved in it. So for sure the proposed system is cost effective than the existing system. Hence the proposed system is economically feasible.

iii. Operational Feasibility

The main problem in the developing the new system is getting acceptance from the user to provide their details to the system. Another problem is co-operation from the users because many users are reluctant to operate new systems. The product being developed is hosted on a webserver so the users are spared from the tedious job of installing software systems. The developed system is usable by range of people who have basic knowledge to surf the internet. Hence the system is operationally feasible

5. Requirements

It is a comprehensive description of the intended purpose and environment for software under development and thus helps in achieving the required specifications for the system

i. Functional Requirements

Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses the functional requirements are captured in use cases. The plan for implementing functional requirements is detailed in the system design. A function is described as a set of inputs, the behavior, and outputs.

ii. Non Functional Requirements

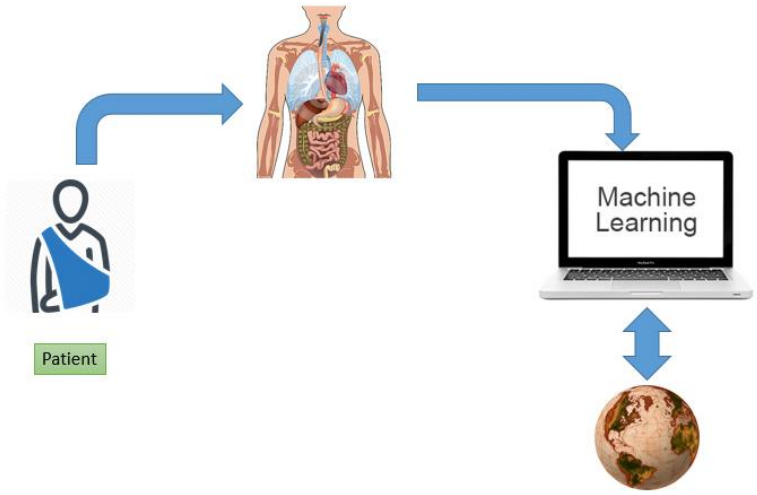
A non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. The plan for implementing non-functional requirements is detailed in the system architecture.

IV. SYSTEM DESIGN

Systems design is the process or art of defining the architecture, components, hardware & software modules, interfaces, and data for a computer system to satisfy specified requirements. This could be seen as the application of system theory to computing and product development. Some overlap with the discipline of systems analysis appears inevitable. Designs tools such as Unified Modelling Language, UML now address some of the issues of computer systems design and interfacing.

A. System Architecture

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.



V. IMPLEMENTATION OF SYSTEM

Implementation is the stage in the project where the theoretical design is turned into a working system. The most critical stage is achieving a successful system and in increasing the reliability of the new system.
It involves careful planning, investigation of the current system and its constraints on implementation, design of methods to achieve the changeover and an evaluation of change over methods.

I. Module Description

Module Description is used to describe the various working modules of the system during implementation phase. The various modules in our system are:

A.Patient Data Gathering

In this module user has to register and the data is collected during the first visit at the outpatient clinics and includes information such as birth date, date of last negative and first positive tests, root cause of infection, and alcohol and drug usage. Data, including clinical data, are collected on a continuous basis every time the patient is seen by the treating physician. This includes information such as the treatment, symptoms of disease, and laboratory results. Information of all kinds is always useful for detecting the disease or the root cause for the patient's problem, hence additional information such as the places visited by the patient in the past week, various allergies that the patient is prone to and various medications that the patient is currently on, is used to determine the exact cause for the patient's distress.

B. Multi Access Control

In this module we are to implement the three different types of account because we have to retrieve three different types of information from the user side and they are listed below. Once the User creates an account, they are to login into their account and request the Job from the Service Provider. Based on the User's request, the Service Provider will process the User requested Job and respond to them.

i. Doctor Account

An account for the doctors ensures that the patient gets the correct care for his disease, although our system is bullet proof, when a patient submits a query for a particular disease he/she is provided with drugs for the ailment, the drugs are prescribed by doctors and hence is highly reliable. The doctor can login into their personal account and check the status of their patients. This includes monitoring the patients for recovery.

ii. Patient Account

The Patients have their own login to monitor the doctor's view on their recovery. This account consists all the details about the patient such as their blood pressure levels, glucose level and also any type of reports they might have undergone during check-up like urine report, X- Ray and MRI scan. It also contains a detailed history of the patient which would prove very useful for the doctor during diagnosis stage.

iii. Administrator Account

The Administrator account is used to primarily have a check on the patient and the doctor as a whole. The administrator has the permission to add doctors into the system who can then be assigned to patients, this feature is particularly useful when specialist doctor from different country have to be consulted for their expert advice. The statistics for the hospital system is available for improvement which can also be used to correct minor problems in the system.

C.Disease based Data Grouping

In this module the big data analyst is going to collect the information about the disease form the insurance server. This can be done once the insurance company accept to expose their customer information to the big data analyst, after which the big data analyst will categorise the data based on the disease. By doing this we can easily get the disease information and list the customers affected by that disease and they can be saved for analysis in future.

D.Blood Report with Pharmacy Analysis

In this module, the patient's health records from various medical tests are stored. Various fields are specified according to the type of medical examination and the body part which is analyzed. Reports like the Blood test, X-Ray etc are stored as the dataset and can be used for the patient disease discovery.

E. Patient Previous History & Personal Data Analysis

The system is designed to represent data that accurately captures the state of the patient at all times. It allows to display entire patient history without the need to track down the patient's previous medical record volume and assists in ensuring data accuracy and legibility. It reduces the chances of data replication as there is only one modifiable file, which means the file is constantly up to date when viewed on a later date and eliminates the issue of lost forms or paperwork. Since all the information is being in a single file, it makes it much more effective when extracting medical data for the examination of possible trends and long term changes in the patient.

F. Big Data Extraction of Useful Information

In this module we implement big data analysis, big data contains a vast amount of data that may or may not contain necessary information. In simple words, the information in the big data are unstructured. So in this module we access the server for big data. The big data analyst get the information and extract the same by the technique of map reducing formation to get useful information which is useful for both disease discovery and to determine the prescription drug for the patient's ailment.

VI. Conclusion And Future Work

In the future we will be guiding the Patient to the Real Best Doctor with respect to the Patients Nearest Location. Feedback of the Previous Patients can also be considered for Identification of Best Doctor. The study explores opportunities for leveraging business intelligence and big data analytics in evidence addressing these gaps. The study presents some of the emerging research areas relating to the use of Big Data Analytics for EBM. In so doing, this study provides a research agenda for health informatics researchers and data scientists to address issues of pressing needs, namely, reducing the cost and improving the cost of healthcare by broadening the practice of evidence based medicine through the applications of business intelligence big data analytics based medicine. The study first describes the steps involved in evidence based medicine and then proceeds to identify current needs and discusses the potential for business intelligence and Big Data analytics in medical and healthcare fields.

References

[1]http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS?order=wbapi_data_value_2010+wbapi_data_value+wbapi_data_value-last&sort=desc, accessed March 08, 2013,
[2] Institute of Medicine, "Shaping the Future for Health", Institute of Medicine, 1999, pp. 1-8.
[3] Bates, D., Cohen, M., Leape , L., Overhage, J., Shabot, M., and Sheridan, T., "Reducing the Frequency of Errors in Medicine Using Information echnology", J Am Med Inform Assoc, 2001, pp. 299–308.
[4] Wells, L., "Role of Information Technology in Evidence Based Medicine : Advantages and Limitations", 2007, pp. 1- 10.
[5] Sackett, D.L., William Rosenberg, J.A., Gray, R., Brian Haynes, and ichardson, W.S., "Evidence Based Medicine: What It Is and What It Isn't", Bmj, 1996, pp. 71-72.
[6] Ibm, "The Value of Analytics in Healthcare: From Insights to Outcomes", in (Editor, 'ed.'^'eds.'): Book The Value of Analytics in Healthcare: From Insights to Outcomes, 2012
[7] West, S., King, V., Carey, T.S., Lohr, K.N., Mckoy, N., Sutton, S.F., and Lux, L., "Systems to Rate the Strength of Scientific Evidence", Agency for Healthcare Research and Quality, 2002,
[8]http://www.bloomberg.com/visual-data/best-andworst/ most-efficient-health-care-countries, accessed Sep 3, 2013,
[9]http://www.cms.gov/Research-Statistics-Data-and- Systems/Statistics-Trends-and- Reports/NationalHealthExpendData/NHE-Fact-Sheet.html, accessed Aug 30, 2013,

[10] Kelley, A.S., Mcgarry, K., Fahle, S., Marshall, S.M., Du, Q., and Skinner, J.S., "Out-of-Pocket Spending in the Last Five Years of Life", J Gen Intern Med, 28(2), 2013, pp. 304-309.
[11] Neubauer, M.A., Hoverman, J.R., Kolodziej, M., Reisman, L., Gruschkus, S.K., Hoang, S., Alva, A.A., Mcarthur, M., Forsyth, M., Rothermel, T., and Beveridge, R.A., "Cost Effectiveness of Evidence-Based Treatment Guidelines for the Treatment of Non-Small-Cell Lung Cancer in the Community Setting", J Oncol Pract, 6(1), 2010, pp. 12-18.
[12]http://www.cancernetwork.com/practice/content/article/1 0165/1821731, accessed Aug 30, 2013, [13] Hurwitz, S.R., Tornetta, P., and Wright, J.G., "How to Read the Literature to Change Your Practice: An Evidence- Based Medicine Approach", The Journal of Bone & Joint Surgery, 88A(8), 2006, pp. 1873-1879